Eye Tracking for Locomotion Prediction in Redirected Walking

Eye Tracking for Locomotion Prediction in Redirected Walking

Markus Zank* Andreas Kunz†

Innovation Center Virtual Reality - IWF - ETH Zurich

ABSTRACT

Model predictive control was shown to be a powerful tool for Redirected Walking when used to plan and select future redirection techniques. However, to use it effectively, a good prediction of the user's future actions is crucial. Traditionally, this prediction is made based on the user's position or current direction of movement. In the area of cognitive sciences however, it was shown that a person's gaze can also be highly indicative of his intention in both selection and navigation tasks.

In this paper, this effect is used the first time to predict a user's locomotion target during goal-directed locomotion in an immersive virtual environment. After discussing the general implications and challenges of using eye tracking for prediction in a locomotion context, we propose a prediction method for a user's intended locomotion target. This approach is then compared with position based approaches in terms of prediction time and accuracy based on data gathered in an experiment.

The results show that, in certain situations, eye tracking allows an earlier prediction compared approaches currently used for redirected walking. However, other recently published prediction methods that are based on the user's position perform almost as well as the eye tracking based approaches presented in this paper.

Keywords: Tracking, locomotion, eye tracking, prediction, redirected walking, virtual reality.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality

1 INTRODUCTION

Walking is our natural way of navigating in our surroundings and for this reason it is also the most natural and intuitive means of navigation in virtual environments. It was shown in a number of studies that users perform better in a variety of tasks the closer the interaction is to the real-world counterpart. In the context of navigation, this means that any interaction using keyboard and mouse is outperformed by navigation methods that allow the user to naturally turn his head to look around. The performance increases further if the user can also physically walk in the virtual environment. This typically requires a large scale tracking space that meets the high latency and precision requirements necessary to avoid cyber-sickness. Howver, the size of the virtual environment will still be limited to the size of this tracking space.

To overcome this limitation, Razzaque et al. introduced the concept of redirected walking [20]. By scaling the user's physical turns in the virtual environment, he can be tricked into turning 180° in the physical world while only turning 90° in the virtual world. In this way, a larger virtual space can be compressed into a limited tracking space.

*e-mail: zank@iwf.mavt.ethz.ch

†e-mail: kunz@iwf.mavt.ethz.ch

IEEE Symposium on 3D User Interfaces 2016
19–23 March, Greenville, SC, USA
978-1-5090-0842-1/16/$31.00 ©2016 IEEE

In the following years, a number of additional techniques for redirection were introduced: curvature gains, which add additional curvature [23], velocity gains, which scale the user's speed in movement direction [7, 27] and resets or reorientations, that force the user to execute specific actions [19, 28]. Usually only one of these techniques was applied, sometimes in combination with an immersion-breaking reset in case the user was about to leave the tracking space. However, to guarantee maximal immersion, all techniques should be used to their fullest potential and exactly in the situations where they are optimal.

To this end, Zmuda et al. [31] introduced the idea of using probabilistic planning to redirected walking. Nescher et al. [14] continued in this direction and formulated the selection of the optimal technique to be applied as a model predictive control problem. This allows applying the minimal amount of redirection required to prevent the user from leaving the tracking space. However, in most cases it is not known in advance what a user will do or where he will go and therefore it is unclear which redirection technique would be best in the long run. In order to solve this problem, a prediction of the user's future path is required.

1.1 Prediction for Redirected Walking

In the past, prediction for redirected walking was done by estimating a single future path which was then transformed to keep the user inside the tracking space [5, 16, 22, 24]. These predictions were made based on the user's current viewing direction, direction of movement, past movement or a combination thereof. The resulting prediction is a single straight line path with the exception of Su et al. [24] who fitted a curved path based on the user's past path. However, a model predictive control planner for redirected walking can consider multiple branching paths in its planning process, and to take full advantage of its potential we require a prediction that provides multiple future paths together with their respective probability.

As Nitzsche et al. [16] already discussed, there are predictions on different time scales and with different knowledge of the environment and models of human behavior. All methods presented and discussed in this paper work on a medium timescale. This timescale is longer than the immediate next step and therefore the predictors are not meant to predict or detect sudden changes in direction such as stopping and turning 90° (however, they should be able to recover and change the prediction immediately after such an event). On the other hand, the prediction will be limited to areas that are currently visible to the user and there are no cognitive or behavioral models included in the prediction. In addition, all methods require discrete target positions the user is expected to walk to. These can either be points of interest or landmarks in the environment or points that are important in the environment's layout such as choke points or junctions.

All predictors used in this paper are formulated to include Bayesian priors and while they need to be determined experimentally for every new environment and can potentially improve the prediction performance, they are not required and can be replaced by a uniform distribution. In this paper, we will use data gathered in one condition of the user experiment to determine the priors and then evaluate the performance of the predictors on data from another condition.

2 Related Work

The prediction of a person's future movements has a wide range of applications and research on this topic has produced a number of different approaches. While prediction of human actions has gained increased importance in the robotics community (e.g., [9, 29]), the requirements and restrictions differ from the ones suitable for an application in virtual reality. In a collision avoidance application, the robot itself carries most of the sensors and has to deal with multiple humans moving around it. Therefore, most methods for prediction have an exocentric perspective on the human and often use camera-based approaches. However, in an immersive virtual environment there is only one person and at least the head position is already tracked in order to control the point of view in the virtual environment which makes this an egocentric problem not usually found in other areas. For this reason, we focus on prediction approaches that already use the egocentric formulation.

Both Zmuda et al. [31] and Nescher et al. [14] used graph representations of their respective virtual environments. This graph representations are predefined and consist of nodes and edges connecting two nodes. On this graph, they used a probabilistic approach to predict the user's future path. For prediction, the edge closest to the user's current position is found and predefined probabilities for the edges connected to the next node are used. However, this does not include any data on how the user actually behaves in the environment over time. In order to use this additional information, a number of prediction approaches based on the user's movement were developed. Su et al. proposed a prediction for telepresence motion-compression by using linear regression of the user's past path to extrapolate his movements into the future [24]. Nescher et al. used a double exponential smoother instead of the regression to smooth the gait-induced head movements and to estimate the intended movement direction [15].

Zank et al. extended this concept for the case in which predefined locomotion targets are available in the environment [30]. By using standard models of human locomotion [1, 2], it is possible to obtain an expected reference path to each of the targets. Once the user starts moving, the observed movement can be compared to the simulated reference trajectories and a prediction can be made. For this, they use standard methods like Dynamic Time Warping and also proposed a new cost-based comparison method that evaluates the amount of "wasted" movement towards each of the potential target points.

However, all these approaches are based on the user's path or location which means that if the paths to two possible targets are identical at this point, there is no way of telling which of the two targets is the correct one. Examples for such situations could be a T-maze, where a long corridor splits up into two opposite paths, or a situation where a person has not yet begun to walk.

However, predicting a person's action is interesting not only in the context of locomotion in virtual reality. It was shown in consumer and cognitive science research that a person's gaze can be highly predictive of future actions in selection tasks. Various researchers found that if people are asked to make a decision between a number of options, there is a significant gaze bias towards the chosen option for a short time period before the decision is announced [3, 17, 25, 26]. It is also known that the gaze direction leads the walking direction by a few seconds [6] and the same holds true for driving [8, 12]. For a comprehensive summary of eye tracking in various tasks including reading, walking and sandwich-making, see Land et al. [11]. In addition, this leading behavior occurs not only for the eyes. It has also been observed that the head direction usually leads the torso and body orientation when walking on a curved path [4].

This prior research leads to the conclusion that a person's gaze could offer benefits over traditional position based approaches when predicting a person's locomotion target. In the following sections, we will discuss the challenges and risks of eye tracking as a prediction method. Then we present a prediction approach using eye tracking and assess its performance in a user experiment.

3 Eye Tracking

It was shown in the past, that our gaze leads future movements when walking or driving. In this context, it seems obvious to use it as a means of predicting future actions. However, other research has shown that gaze is also directed to regions that contain information relevant to the current task [18, 21, 25] which is in line with observations made in the driving context. We assume that since driving is usually the primary task, gaze is mostly controlled by the requirements of steering and therefore a lot of time is spent looking at the intended driving target, especially in the often artificial experimental context this data was gathered in. However, for data gathered on the road, the gaze will already be distributed between steering along the road, watching other traffic participants, and the scenery [10].

But in the context of real walking in virtual environments, walking around might not even be the primary task. Consider for example a virtual art gallery, it is very likely that a person's gaze will be directed towards the paintings, since this is usually the primary reason for being in an art gallery. However, it is reasonable to assume that the visitor will not walk to each of them. But he will still look at most of them, at least to determine if it is worth to walk there. A similar situation occurs for a search task, there will be a lot of scanning the environment and a decision for a new locomotion target will only be made if there is new information visible. In these situations, it is possible that eye tracking alone cannot provide sufficient data for a stable prediction, but it might be able to support an unsure decision based on a position based prediction.

At this point, we have to make a distinction between locomotion and wayfinding. Following the definitions by Montello [13], locomotion is the physical movement of the body through the environment, whereas wayfinding is the cognitive task of planning a future movement. Together, they are involved in navigation which he defines as "goal-directed movement through the environment". In the cases described above, it becomes difficult to distinguish eye movements used for wayfinding from those used without locomotion intention. However, we hypothesize that there is a point where we need our vision for locomotion in order to make a certain turn or avoid obstacles, similar to the behavior observed for driving a car. At this point, eye tracking should be able to provide evidence for a prediction, but it is unclear how much time can be gained compared to traditional approaches.

4 Methods

The prediction method presented in the following section assumes that there is a set of target points representing a person's possible future targets. Automatically detecting these points from the environments is beyond the scope of this paper and it is assumed that the location of these points is known.

The goal is to estimate the probability $P(\tau)$, which is the probability of τ being the intended locomotion target, for each target $\tau \in \mathscr{T}$ given the user's current behavior. $\mathscr{T}$ is the set of all available targets and in this paper we assume that it contains all possible targets (meaning there are no targets we don't know about). In the following, we present a method to use eye tracking data to achieve this goal and compare it with prediction approaches based on the user's position.

It can be assumed that there is a gaze bias towards a person's intended locomotion target τ. Additionally, it can be assumed that there is a relation between the user's location in the environment R and the point in the environment he is looking at G given a certain target. For this reason, we propose to use a probability distribution for G that is conditional on the user's location and his locomotion

target. Equation (1) follows from Bayes' theorem and describes the probability of a certain target being the one the user intends to walk to $P(\tau_i|R,G)$ as a function of $P(R,G|\tau_i)$ which has to be determined experimentally. However, it might be possible in the future to automatically generate them based on the virtual environment, if a bottom-up model based on the environment's local geometry and saliency is viable. If this is not the case, a data driven approach could be more promising. $P(\tau)$, on the other hand, contains a prior probability and allows bringing in prior knowledge about the probabilities of all targets. However, these values are highly environment-, task- and knowledge-dependent which makes them difficult to determine but they can potentially be a powerful tool for highly stereotyped or scripted tasks. This could for example be the case in a training scenario where the correct sequence of tasks is known and the user is supposed to follow it.

$$P(\tau_i|R,G) = \frac{P(R,G|\tau_i)\cdot P(\tau_i)}{\sum_{\forall \tau_j \in \mathscr{T}} P(R,G|\tau_j)\cdot P(\tau_j)} \quad (1)$$

Please note that even though $P(R,G|\tau)$ is formulated as a multimodal probability distribution, it merely serves as a fully complete example in this section, while the formulations evaluated in this paper will be discussed in Section 6.1. G and R are points in 3D space, but for a virtual environment that does not allow vertical movement, it is sufficient to use a 2D position for both. The probability is realized as a discrete probability distribution, where the resolution can be chosen according to the requirements and the amount of available data.

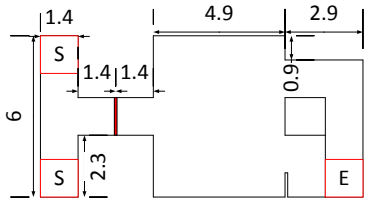
A second approach is to look at differences in the temporal distribution of G. The hypothesis is that there is a significant bias towards the selected target similar to the one observed by Wiener et al. [25]. From their work and other work on eye tracking during decisions, we hypothesize that there are two areas or time intervals in which the bias could occur: First, when the decision is made, and second when gaze is used for steering. The first occurrence should be influenced by prior knowledge of the environment and task-specific circumstances, while the second one should depend more on the local geometry (sharp turn, narrow passage, door, etc.). Since both aspects are directly related to the environment's geometry, we will consider the problem more from a spatial rather than a temporal perspective and will therefore focus on distances and zones in space rather than time intervals in the evaluation.

5 Experiment

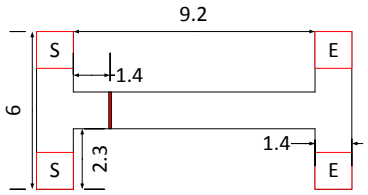
5.1 General Considerations

In order to gather data to determine $P(R,G|\tau)$ and to compare eye tracking based prediction with position based prediction, a user study was conducted. Since prior knowledge and task can play a major role for eye movements, it was decided to conduct a user experiment without a task that could strongly influence eye movements. Such a task could be a search task where a user would naturally use their eyes to look for whatever they need to find.

To be able to evaluate the eye tracking and path data, it is important to know the user's intended target. Other work that demonstrated the potential of eye tracking for prediction often used selection tasks, where the selection can be performed almost instantaneously after the user's decision. However, for walking there is an inherent delay between the choice of a target and the time when the target is reached. It is therefore possible that multiple decisions occurred during the walking process, for example if a person can't decide between two targets or changes his mind. In such a case the prediction should show an alternating prediction. However, without any additional insight on the participant's cognitive process (e.g., through a think-aloud protocol), we would assume poor prediction performance, even though it was actually correct. Therefore, it is necessary for the user to know where he wants to go to be able to properly determine the predictors performance in a post-hoc evaluation. For this, it was decided to give the participants clear instructions on what to do and not let them explore an unknown environment or make too many decisions on their own.



(a) Y-room, left condition. The right condition is mirrored.



(b) T-maze, all conditions

Figure 1: The main layouts. The user enters from one of the two elevators on the left (marked with S), approaches the door with the instruction (red line) and walks to one of the elevators on the right. The elevators take him to the respective elevator (marked with E) on the next floor. All distances are in meters.

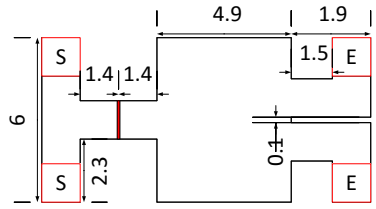


Figure 2: Y-room, free choice condition. Due to space restrictions, the layout for the free choice condition in the Y-room is slightly different from the left and right conditions. All distances are in meters.

5.2 Design

In the final design, the users were placed in a virtual environment of a multi-story building and instructed to reach the top floor by following arrows located in the environment and using a series of elevators for vertical movement. Every condition in the experiment was located on a single floor, each of which had two exits or elevators leading to the next floor. In this way, any number of conditions and repetitions can be done using the maximum available tracking space without redirection or interaction from the experimenter, while providing enough narrative to the user to behave naturally.

As already discussed by Zmuda et al. [31], we can expect different behavior from people in a narrow maze-like environment in comparison to wide, open spaces. Prediction is also potentially easier in the second case because the traveled path will diverge earlier compared to a narrow maze, since humans generally follow short and smooths trajectories to their target. For this reason, two main

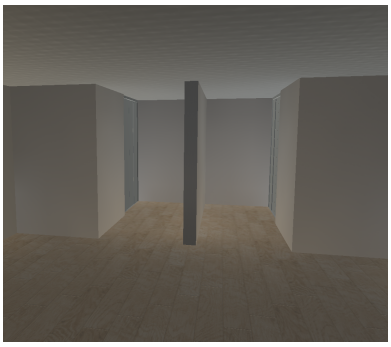
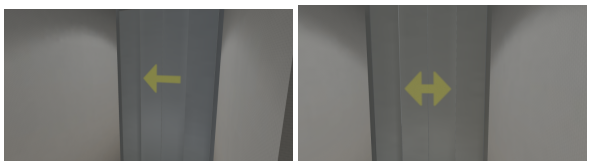


Figure 3: View in the Y-room "free choice" condition from the starting door towards the exits.



(a) Go left (b) Free choice

Figure 4: Instruction symbols. They appear as holograms on the currently closed door.

layout types were included in the study: A room with two exits side-by-side on the opposite wall (Y-room, Figure 1a) and a corridor with two exits to the left and right (T-maze, Figure 1b). In both cases, the layouts have two exits that are located symmetrically with respect to the room's main axis. To avoid any changes in visibility while the participant is moving through the environment, a door was added at the beginning of the floor. This door is initially closed and only opens once the user is standing in front of it for two seconds.

Each layout was paired with one of three instructions: "Go left", "go right" or "free choice". The "go left" and "go right" conditions simulate a participant with knowledge of the environment and a clear target in mind. This can be either task-related, or because the participant wants to go to a certain location in the environments. In the "free choice" condition, the participant does not have a predefined target, this can either represent a search task, an unknown environment, or a situation where both decisions lead to the intended target. For space reasons the layout for the "free choice" Y-room has a smaller distance between the exits (see Figures 2 and 3), the layout for the T-maze is identical for all conditions. Participants were instructed of their task on the individual floors with arrows that appear as holograms on the closed door at the beginning of the floor (see Figure 4). Once the door opened, the icon disappeared. This allows a comparison with the other two conditions to see how the gaze-patterns changes from the forced conditions.The left and right conditions appeared three times each, together with two free choice conditions for each of the room layouts.

In addition, there were a number of "museum" floors. Here, the users were shown sets of three paintings, symbols or numbers and had to indicate which of them does not fit with the others by standing in a circle in front of it (see Figure 5 for an example). This task was included to keep the users engaged and also to distract from the experiment's actual purpose. Therefore, the sets were deliberately designed to be non-trivial with potentially more than one answer that could be argued for. Such a "museum" floor was inserted after every other study floor. This results in a total of 25 floors with 16

Figure 5: View from the entrance to the three symbols on the 12^{th} floor

study layouts, seven "museum" floors, a start and an end floor.

Prior to the experiment, the participants had the opportunity to walk around in a virtual environment of a small apartment in order to familiarize themselves with the system. Afterwards, they were instructed about their task for the experiment with a short example. This example used the same geometries and arrows, but a different "museum" floor. After they finished the experiment, they were offered the opportunity to try another virtual environment.

5.3 Setup

The setup (depicted in Figure 6) consists of an Oculus DK2 head-mounted display[1] with an integrated SMI eye tracker[2]. The tracking system is an Intersense IS-1200 attached to the HMD providing 6 DOF position tracking at 180 Hz. The system is powered by a backpack-mounted laptop and the virtual environment runs in Unity3D[3]. The environment was optimized to run constantly at the display's maximum framerate (75 Hz).

The eye tracker was calibrated using a 5 point calibration and runs at 60 Hz. The gaze vector given by it was intersected with the geometry in real time and both the original vector and the resulting 3D position were recorded. The data from the position tracker was recorded at 180 Hz, but was subsampled to match the eye tracker's update rate.

Additionally, the users wore headphones and heard music and sounds from the doors and elevators in the environment in order to dampen the sounds in the real environment.

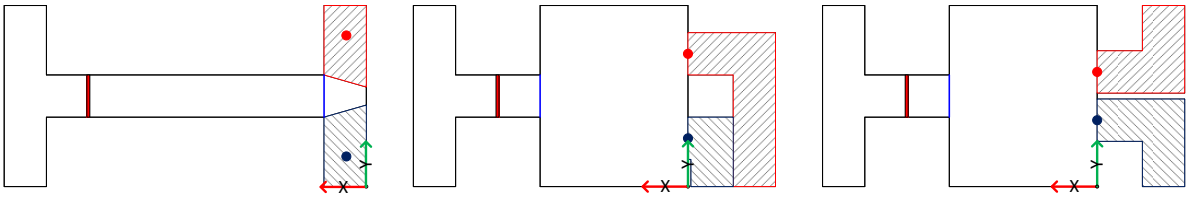
6 Results

14 participants were recruited among the student body. The average age was 25 ($\pm$ 4.1) years and average height was 177 ($\pm$ 7) cm. The data was then cut into individual paths on each floors, resulting in 224 study-floors. The beginning of each path is the first eye-tracking data point that is beyond the initial doors. The path ends when the participant enters one of the target zones. The data from the start and end floors as well as the "museum" floors was recorded but not used in this evaluation. 24 floors were excluded from the evaluation, because users went to the wrong elevator either because they forgot the instruction or they were curious to see what is there.

[1] http://www.oculusvr.com

[2] http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/eye-tracking-hmd-upgrade.html

[3] http://unity3d.com



(a) T-maze, all conditions (b) Y-room, left (c) Y-room, free choice

Figure 7: Target zones (hatched area) and target locations (red and blue circles) for all conditions. The red vertical lines represent the doors at the beginning of the environment, the blue vertical lines mark the location where the environment opens up to the side (see Figures 12-15).

Figure 6: Virtual reality simulator setup

6.1 Eye Tracking Realization

For the evaluation of the presented user experiment, three realizations of the eye tracking concept in Section 3 are used. As mentioned before the probability that are conditional on τ are discrete probability distribution and their resolution is limited by the amount of available data. For the eye tracking data, we propose a discretization that only allows one bin per target plus one additional bin that is not associated with any target. We will refer to these areas as target zones $\mathscr{G}$, since they correspond to a physical area around the target points. While this is a very rough discretization, it is very easy to determine them automatically for a new environment given the targets' locations. Figure 7 defines the target zones used defined the layouts used for the experiment.

The first version is a purely eye tracking based approach that does not take the user's position in space or changes over time into account. From the gathered data , we calculate $P(G \in \mathscr{G}_i|\tau_i)$ (see Table 1) and determine the probability as defined in equation (2).

$$P(\tau_i|G) = \frac{P(G \in \mathscr{G}_j|\tau_i) \cdot P(\tau_i)}{\sum_{\forall \tau_j \in \mathscr{T}} P(G \in \mathscr{G}_j|\tau_j) \cdot P(\tau_j)} \tag{2}$$

The second method that was considered takes the user's position in the room and the eye tracking position into account. In the general case, the user's position in 2D space should be considered. However, because the layouts used in this study have one main direction of travel, it is acceptable to only consider the user's position in this direction (x-axis). This allows capturing the changes in the gaze distribution when the user approaches the decision point. The probability is defined in equation (3).

$$P(\tau_i|G,R_x) = \frac{P(G \in \mathscr{G}_j|\mathscr{R}_i \leq R_x < \mathscr{R}_{i+1},\tau) \cdot P(\tau_i)}{\sum_{\forall \tau_j \in \mathscr{T}} P(G \in \mathscr{G}_j|\mathscr{R}_i \leq R_x < \mathscr{R}_{i+1},\tau_j) \cdot P(\tau_j)} \tag{3}$$
$$\text{for } \mathscr{R}_i = f \cdot i, \mathscr{R}_i \in [0,20]$$

The range $[0,20]$ is chosen to completely cover the layouts used while the resolution f can be chosen depended on the available data. The methods proposed so far cannot capture behavior over time such as a user looking at a single target for a long time. To account for this, we propose a combination of the single frame prediction defined in equations (2) and (3) over a defined time period T, where t_0 is the time of the prediction. By tuning T it is possible to trade responsiveness and the maximal confidence. For example a predictor with a very small T will react quickly if the user looks somewhere else, but there is also a upper limit for $P(\tau_i,t_0)_T$. A larger T allows $P(\tau_i,t_0)_T$ to go to 1 if the user looks at a target long enough, but the predictor is not as responsive.

$$P(\tau_i,t_0)_T = \frac{\prod_{t=t_0}^{t_0-T} P(\tau_i,t)}{\sum_{\forall \tau_j \in \mathscr{T}} \prod_{t=t_0}^{t_0-T} P(\tau_j,t)} \tag{4}$$

The next method is based on the gaze bias towards the chosen target that was expected based on the literature. The user is expected to look at the intended target for a longer period overall, but not necessarily in the last time. Instead of using the previous method with a very large T, we consider the expected distribution of stays in different target zones. $T_{spent}(i)$ is the total time the user was looking at G_i so far. $P(T \geq T_{spent}(i)|\tau_i)$ is the probability that a person would look at G_i for at least T_{spent} given that i is the intended target.

$$P(\tau_i|T_{spent}) = \frac{P(T \geq T_{spent}(i)|\tau_i) \cdot P(\tau_i)}{\sum_{\forall \tau_j \in \mathscr{T}} P(T \geq T_{spent}(j)|\tau_j) \cdot P(\tau_j)} \tag{5}$$

The main metric for comparing prediction methods is the time between the time a reliable prediction and the time the predicted action (in this case turn left or right) occurs. However, since this time depends on the user's walking speed, the predictions are compared based on the distance between the point where a reliable prediction can be made and the wall opposite of the door through which the user entered the room projected on the room's x-axis. The axes and origins for the respective layouts are defined in Figure 7.

6.2 Statistical Analysis

In the following section, the observed user behavior is analyzed statistically. In the "free choice" condition, participants chose the left path in the T-maze condition in 14 cases, while the right path was

Condition	Intended Target	Other Target	Off Target
T-maze Left	0.58	0.19	0.23
T-maze Right	0.50	0.23	0.27
T-maze Free	0.46	0.28	0.26
Y-maze Left	0.65	0.12	0.23
Y-maze Right	0.67	0.11	0.22
Y-maze Free	0.48	0.12	0.40

Table 1: Relative overall stay time is in a certain target-zone

chosen also chosen 14 times. For the Y-room the left path was chosen 12 times, the right path was chosen 15 times. 4 participants chose the same option in both T-maze conditions while 7 participants chose the same option in both Y-room conditions. Since there is practically no task in this study, it is possible that there is an influence of the previous floor's layout.

Figure 8 and Figure 9 show the duration of the stays in different target-zones. In this context a "stay" is considered to be the period during which the participant looks at one specific target zone. There can be multiple fixations within one zone, meaning that multiple points can be focused on within the same target zone, but it still counts as one stay. Both the number of stays and the duration of the individual stay is significantly higher for the zone of the intended target ($p < 0.01$). Table 1 shows the relative stay time in the zone they eventually entered, in the other zone and in neither of the zones for the different conditions and layouts. For the "free choice" conditions this was assigned based on the final decision. Even when people were free to choose, there is a bias towards the target-zone that was entered at the end.

6.3 Single Sample Evaluation

In the following, the data is analyzed looking at only a single data point or a pair of data points from position and eye tracker. Figure 11a shows all the sampled user positions in the xy-plane. Figure 11b shows the gaze direction (as defined in equation (6) and Figure 10) along the users' paths through the environment. Every line corresponds the path traveled by a single participant on one floor.

$$\varphi = \tan^{-1}\left(\frac{G_y - R_y}{R_x - G_x}\right) \tag{6}$$

In general, both position and gaze direction show a highly stereotypical behavior without any fundamental differences in behavior between people. However, as expected there are some cases when people look around or scan the alternative corridor before looking in the intended walking direction causing the outliers in Figure 11b.

For the Y-room condition, the behavior is different as shown in Figure 12. Due to the wide space available, the users' paths diverge around 5 meters in front of the opposite wall which is exactly at the point the room widens. This aligns nicely with the principle underlying many locomotion models that humans try to keep the path traveled as smooth as possible. Since φ takes the y-position into account when calculating the angle, the gaze direction differs during the steering motion around 4-5 meters in front of the opposite wall and goes back to zero afterwards.

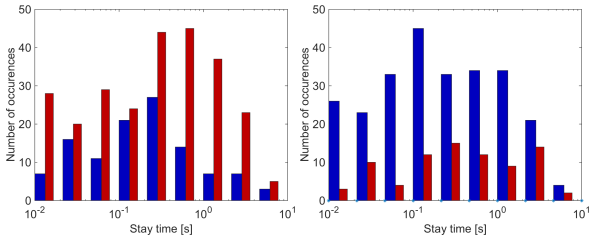
6.4 Position Based Prediction

In order to determine if and to what extent an eye tracking allows an earlier prediction compared to approaches based purely on the user's position, the approach previously presented in this paper is compared with previously published methods.

The compared approaches are the one used by Nescher et al. [14] for redirected walking, the two cost-based approaches and one approach using Dynamic Time Warping (DTW) by Zank et al. [30]. The approach by Nescher et al. uses a graph representation of the environment consisting of straight line segments and arc segments with 0.8m radius. The prediction is done by finding the graph segment closest to the current position and following the graph along the direction of movement. When a node is reached, predefined probabilities are used for all possibilities.

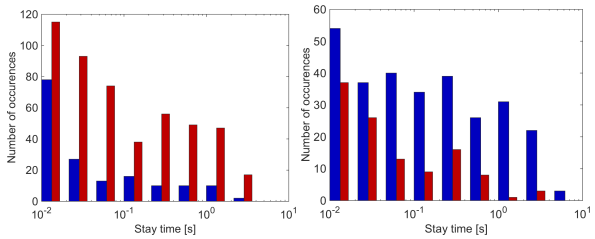
The three predictors by Zank et al. are centered around the idea that human locomotion trajectories are optimal to some degree, meaning that humans usually walk on smooth and short paths to their target. They propose to use a model of human locomotion to generate a set of reference trajectories to all possible targets the moment the user starts to walk and then compare the observed trajectory to them while the user walks. In the following evaluation the models by Arechavaleta et al. [1] and Fink et al. [2] are used to generate these reference trajectories. In addition, they presented different schemes for comparing the observed path to the reference trajectories, the most promising ones are based on Dynamic Time Warping (DTW) and the cost-based method. Here, they use a cost function to calculate the amount of movement that was "wasted" if the user wanted to reach a specific target. Based on the difference in "wasted" movement towards different targets, they calculate a probability distribution. The combination of the path model by Arechavaleta et al. with the cost-based predictor ("Arechavaleta, cost") as well as the model by Fink et al. with both the cost-based ("Fink, cost") and the Dynamic Time Warping ("Fink, DTW") predictors are included in the evaluation.

The approaches by Nescher et al. and Zank et al. need distinct points in the environment as target points, they are manually defined as depicted in Figure 7. The predictors by Zank et al. use the first position sample on the path as a starting point, the approach by Nescher et al. uses an additional node in the middle of the door. For the following evaluation, both targets were considered by the predictors. Even for the "go left" and "go right" condition, the predictors were not aware that the user had no choice. Otherwise,



(a) Left (b) Right

Figure 8: Distribution of the stay time between the left target-zone (red) and the right target-zone (blue) for the conditions T-maze go left (a), and T-maze go right (b)



(a) Left (b) Right

Figure 9: Distribution of the stay time between the left target-zone (red) and the right target-zone (blue) for the conditions Y-maze go left (a), and Y-maze go right (b)

T / Y	Arech., cost		Fink, cost		Fink, DTW		Nescher Graph		$P(\tau\|G)_{0.25}$		$P(\tau\|T_{spent})$	
More than 50% classified at [m] (x-axis)	9.6	6.3	9.4	6.6	-	3.9	1.1	4.9	8.2	6.5	3.2	4.1
Avg. $\frac{\#\text{ correct}}{\#\text{ class}}$	0.48	0.75	0.59	0.76	1.00	0.92	0.99	0.88	0.69	0.78	0.77	0.79
Max. $\frac{\#\text{ correct}}{\#\text{ class}}$	0.64	0.93	0.99	0.94	1.00	1.00	1.00	1.00	0.86	0.97	0.88	1.00
Mean overall performance	0.48	0.73	0.59	0.78	0.50	0.71	0.51	0.81	0.62	0.78	0.58	0.70
Std dev. overall performance	0.36	0.24	0.17	0.20	0.01	0.09	0.00	0.08	0.22	0.22	0.10	0.14

Table 2: Characteristic performance for predictors for the forced T-maze condition in the first column under each method and the forced Y-room condition in the second column. The two best performing and earliest predictors are highlighted for each layout.

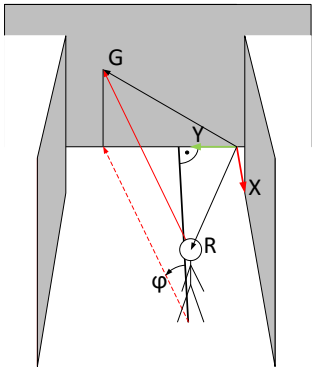


Figure 10: Projection of the gaze vector on the ground plane. φ is the angle between the gaze direction and the room's main axis.

all predictors used in the evaluation would give a probability of 1 for the correct target.

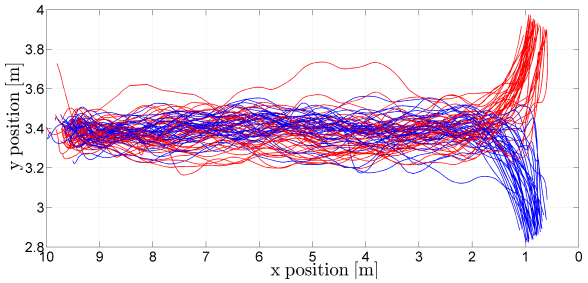
6.5 Comparison

The eye tracking based predictors were evaluated in a leave-one-out cross-validation scheme. The prediction for each user was done based on the conditional probabilities calculated with the data from the remaining participants. The combination of measurements over a certain time period as defined in equation (4) was evaluated for $T = (0,1)$. There is an increase performance for larger T, but the increase levels off for $T > 0.25$ and since this also increases latency, $T = 0.25$ was chosen for the following evaluation.

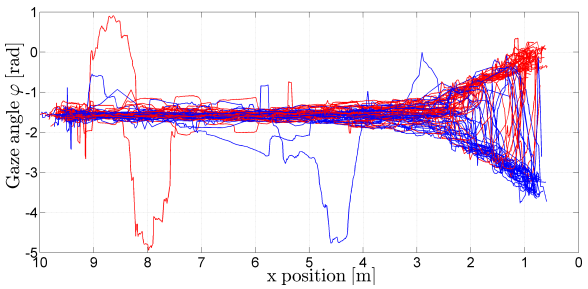
Both the $P(\tau|G,R_x)_{0.25}$ and $P(\tau|G)_{0.25}$ predictors (equations (2) and (3)) were evaluated, but the difference in performance was small and not significant ($p = 0.72$). As a result, only $P(\tau|G)_{0.25}$ will be included in detail in the following evaluation.

To compare the different prediction methods, we assume that a prediction is considered reliable when the maximum probability reaches 66%. Then we look at the user's position along the x-axis when a reliable prediction is made. Figures 13 and 14 show the ratio of correct, incorrect and undecided samples for the the T-maze and Y-room conditions against the user's current position along the x-axis. This visualizes how the prediction develops as the users progress through the respective environments.

Table 2 summarizes the predictors' characteristic performance figures, namely the largest distance from the opposite wall at which more than 50% of all users' paths were classified (correctly or in-

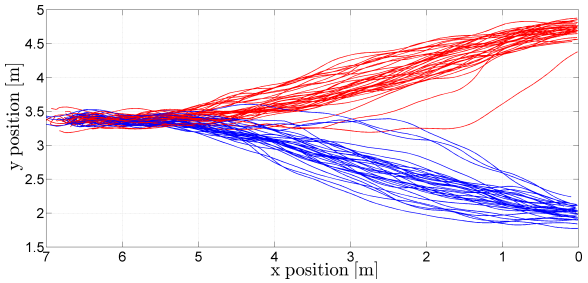


(a) 2D position of the participant

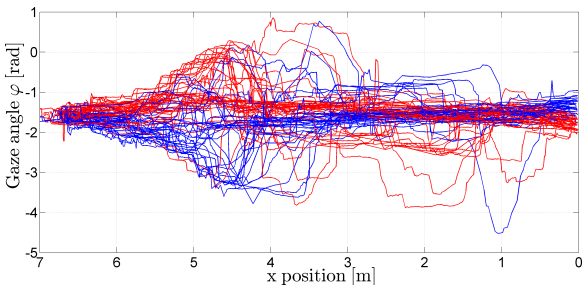


(b) Gaze direction and x-position of all participants according to the definition in Figure 10

Figure 11: Position and gaze angle for the "go left" (red) and "go right" (blue) conditions for the T-maze layout.



(a) 2D position of the participant



(b) Gaze direction and x-position of all participants according to the definition in Figure 10

Figure 12: Position and gaze angle for the "go left" (red) and "go right" (blue) conditions for the Y-room layout.

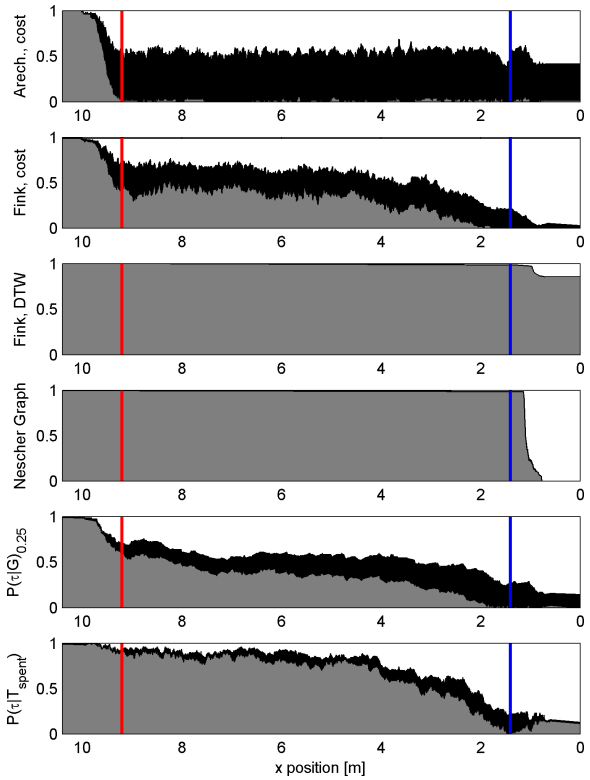


Figure 13: Distribution of prediction performance along the x-axis for the selected predictors for the T-maze layout. The plots show the ratio of correct (white), incorrect (black) and undecided (gray) data points along the room's x-axis as well as the location of the initial door (red vertical line) and the opening into a wider space (blue vertical line).

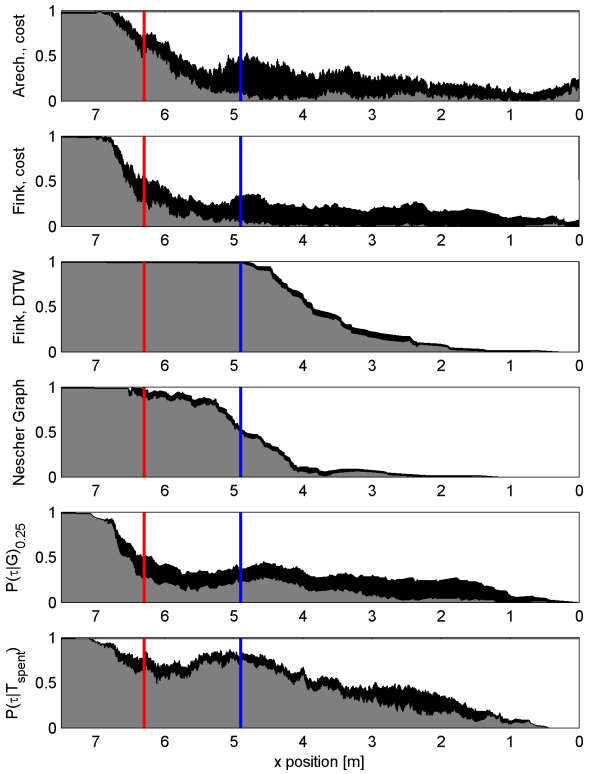


Figure 14: Distribution of prediction performance along the x-axis for the selected predictors for the Y-room layout. The plots show the ratio of correct (white), incorrect (black) and undecided (gray) data points along the room's x-axis as well as the location of the initial door (red vertical line) and the opening into a wider space (blue vertical line).

correctly) and the maximum and average ratio between the number correctly classified samples to the overall number of classified samples. The performance of a single path is defined to be the average of the probability of the correct target equation (7) where L is the overall length of the path. The predictor's overall performance is the average performance of all recorded paths for the "go left" and "go right" conditions of one layout.

$$E = \frac{1}{L}\int_0^L P(\tau_{correct})dx \tag{7}$$

For the following evaluation, the performance of all individual paths is compared using ANOVA. Results are considered to be significant if $p < 0.05$. Since the predictors were tested in a pairwise fashion and the results were Bonferroni-corrected for multiple comparisons. Overall, $P(\tau|G)_{0.25}$ has the highest performance and is significantly better than the predictors "Arechevaleta, cost" ($p < 0.01$), "Fink, DTW" ($p < 0.01$) and $P(\tau|T_{spent})$ ($p = 0.037$). The "Fink, cost" predictor is also significantly better than "Fink, DTW" ($p < 0.01$), while "Nescher Graph" is not significantly different from any other predictor.

For the T-maze condition, $P(\tau|G)_{0.25}$ also significantly better than "Arechevaleta, cost" ($p = 0.049$), "Fink, DTW" ($p < 0.01$), "Nescher Graph" ($p < 0.01$) and $P(\tau|T_{spent})$ ($p < 0.01$). The "Fink, cost" and the $P(\tau|T_{spent})$ predictor are also significantly better than "Fink, DTW" and "Nescher Graph" ($p < 0.01$ for all combinations).

For the Y-room condition, $P(\tau|G)_{0.25}$ is significantly better than "Fink, cost" ($p = 0.034$) and the $P(\tau|T_{spent})$ ($p = 0.027$). The "Nescher Graph" predictor is also better than both of them ($p < 0.01$).

To compare which predictor offers an early prediction, we look at the point where a decision has been made for at least 50% of all users. This corresponds to a position along the environments x-axis and is stated in Table 2. The "Fink, DTW" predictor is excluded for the T-maze, because it never gives a prediction for the majority of users. It can be seen that "Arechavaleta, cost" is the earliest predictor for the T-maze condition, "Fink, cost" is only 0.2 meters later. The best eye tracking based predictor is $P(\tau|G)_{0.25}$ which is 1.4 meters behind. For the Y-room condition, "Fink, cost" offers the earliest prediction, followed by $P(\tau|G)_{0.25}$ and "Arechavaleta, Cost". The remaining predictors are significantly later.

The initial hypothesis was that there should be two zones where eye tracking should work best. First, when the environment is initially seen and the user orients himself and a second time when gaze is used for steering. The vertical red lines in Figures 13 and 14 show the location of the door whereas the point at which the room opens up to the side is marked with a blue line (also see Figure 7).

For $P(\tau|G)_{0.25}$, the advantages of the eye tracking prediction can be seen. For both conditions, there is a number of path that can be predicted correctly even before the participant has passed the initial door (red vertical line). The cost-based predictors show a similar behavior. The predictor using the closest point on the graph (Nescher Graph) and the "Fink, DTW" predictor are not able to give a prediction that early.

For the Y-room conditions on the other hand, the "Fink, DTW" predictor and graph based approach both perform well with a very small number of wrong predictions, but they are both later than the

	T maze	Y room	overall
Arech., cost	0.46 (-0.01)	0.66 (-0.07)	0.56 (-0.04)
Fink, Cost	0.55 (-0.04)	0.71 (-0.07)	0.63 (-0.05)
Fink, DTW	0.50 (-0.00)	0.61 (-0.11)	0.55 (-0.05)
Nescher Graph	0.51 (-0.00)	0.69 (-0.12)	0.60 (-0.06)
$P(\tau\|G)_{0.25}$	0.52 (-0.10)	0.74 (-0.06)	0.63 (-0.08)
$P(\tau\|T_{spent})$	0.55 (-0.02)	0.67 (-0.03)	0.61 (-0.03)

Table 3: Performance of the predictors for the "free choice" conditions. In parentheses is the difference to the "forced" condition. Significant changes are highlighted.

	T maze	Y room
Arech., cost	9.54 (-0.02)	5.16 (-1.18)
Fink, Cost	9.01 (-0.36)	6.55 (-0.02)
Fink, DTW	- (-)	3.29 (-0.60)
Nescher Graph	1.05 (-0.05)	4.06 (-0.83)
$P(\tau\|G)_{0.25}$	4.93 (-3.24)	6.59 (0.10)
$P(\tau\|T_{spent})$	3.28 (0.05)	6.55 (2.42)

Table 4: Difference for more than 50% classified [m]. In parentheses is the difference to the "forced" condition. Notable changes are highlighted.

cost-based and the $P(\tau|G)_{0.25}$ predictor. For the T-maze condition, they are only capable of providing a good prediction after the participant has walked into the side corridors (marked by the vertical blue line). This is a serious limitation, on the other hand they have the smallest number of wrong classifications for both conditions which could be very useful when combining multiple predictors.

6.6 Free Choice Conditions

In the following section, the prediction is evaluated for the "free choice" condition. The Bayesian priors are determined based on the position and eye tracking data from the "go left" and "go right" conditions and resulting probability distributions.

As in the previous section, the conditions are compared based on their performance as defined in equation (7). For every predictor, the performance for the free "free choice" condition is compared with the performance of the same predictor for the forced condition. Table 3 summarizes the performance of the predictors and shows the difference when compared to the "forced" condition.

While the two cost-based predictors show no significant difference between the free and forced conditions, the DTW and graph-based predictor shows a difference for the Y-room condition (both $p < 0.01$). From the eye tracking predictors only $P(\tau|G)_{0.25}$ shows a significantly worse performance in the T-maze condition ($p = 0.03$), which is mainly caused by a worse performance in the beginning. There is no difference for the Y-room and the $P(\tau|T_{spent})$ predictor. Looking at the change in prediction distance shown in table 4, there is little change with two exceptions. $P(\tau|G)_{0.25}$ is 3.24 meters later than in the "forced" condition whereas $P(\tau|T_{spent})$ is 2.42 meters earlier than in the "forced" condition

7 Discussion

In general, eye tracking based prediction preforms well compared to recent position based approaches using path models. However, both the Y-room and T-maze conditions the "cost" predictors are capable of giving a prediction earlier than the eye tracking based predictors, but they also make more wrong predictions. On the other hand, compared to the "Fink, DTW" and "Nescher, Graph" predictors which have a lowest number of wrong predictions, eye tracking allows a prediction much earlier.

Looking at the changes in the prediction's certainty (number of either correct or incorrect samples to number of undecided samples) in relation to key geometrical points in Figures 13 and 14, we can distinguish a number different behaviors. In some cases, there is an initial increase immediately after the door opens (marked by the red vertical line). This is the case for both eye tracking and both cost-based predictors. The second increase happens when the geometry opens up and allows paths to diverge (blue vertical line). Some predictors (like "Nescher Graph" for the Y-room and "Fink, cost" for the T-maze) are able to detect this development well before the user actually reaches this point while others ("Fink, DTW" for both conditions) have an increase in certainty immediately afterwards, resulting in a steep slope in the classified to unclassified sample ratio. It is noteworthy that the "Arechevaleta, cost" predictor behaves more like the eye tracking based predictors in this respect. It shows an initial increase in prediction certainty but than levels off and is not capable of taking advantage of the diverging paths in the Y-room condition. While it still exhibits good performance in the Y-room, for the T-maze this results in a high certainty, high error output.

For the T-maze condition, the $P(\tau|G)_{0.25}$ predictor and the cost-based predictors both perform significantly better than the the approach by Nescher et al. However, in the Y-room condition, this approach performs very well and while it can not predict as early as the eye tracking and cost-based approaches, it has an exceptionally good right-to-wrong classification ratio. This suggests that different predictors perform well under certain circumstances and combining them dependent on the environment or in a voting scheme should be considered in the future.

However, the problems with the independence of gaze and movement intention discussed in the beginning might be a problem for the use of eye tracking based prediction. Even in the simple and clean environment used in this experiment, there was a high variance between participants in the eye tracking data especially when compared to the highly stereotypical path behavior. This could also be the reason that the single sample eye tracking predictors performed poorly and only when averaging over time, the prediction was stable enough for prediction. This problem can be expected to intensify in a more visually distractive or busy environment that will be found in a real world application for Redirected Walking. In this case, it will probably be necessary to model the gaze behavior to differentiate the baseline gaze behavior from subtle changes dependent on the intended target, similar to the path models and comparison in [30].

8 Conclusions and Future Work

The results show that eye tracking is indeed capable of providing an early prediction of a person's intended locomotion target. While the performance in the Y-room condition with a lot of open space is comparable to the top position based predictors, eye tracking offers a clear advantage for the narrow T-maze layout where it offers a prediction as early as the "cost" based predictors, but with a much higher number of correct predictions. It is also capable of providing a prediction 6.6 meters before the previously used graph-based approach by Nescher et al. can. However, it never achieves their 100% correct result, mainly because a number of participants peeked in the opposite direction while turning at the end of the T-maze. This is one of the main problems with using eye tracking data, even in the very clean and empty environment used in this study, some people looked around to orient themselves and explore the environment resulting in a relatively high percentage of wrong predictions.

The second problem with the current formulation of the eye tracking based predictors is the use of conditional probabilities. However, since the predictions conditional on the user's position did not perform better than the ones depending only on the gaze, the only remaining dependency is between gaze and target. It is likely that this is influenced by the number and visibility of the

available targets, but the data in Table 1 shows a very strong bias towards the intended target. It might therefore be possible to find a environment-independent approximation in which case all conditional probabilities can be determined in advance. In addition, it should be possible to learn the conditional probabilities over time by gathering data from users in the respective environment.

To increase the prediction performance in general, a majority or confidence vote should be added in the future to be able use multiple predictors simultaneously and combine their predictions. This could be done by either using a simple voting scheme or by using different predictors in different zones along the user's path.

While the eye tracking based prediction has demonstrated good performance in this experiment, it is necessary to further evaluate its performance in more visually cluttered environments. If performance can be confirmed in realistic scenarios, eye tracking can offer egocentric prediction of locomotion in environments where position tracking is not available such as applications like pedestrian navigation.

In the future, the benefit of the improved prediction on Redirected Walking needs to be evaluated as well. The number of resets required per time should be reduced with improved prediction, however it is not immediately clear how much improvement can be achieved with a prediction on this time scale. In order to use the proposed prediction more easily, it is also necessary to extract the target points, gaze zones, and paths automatically from the virtual environment.

Acknowledgements

The authors wish to thank the Swiss National Science Foundation, project number 205121 153243, for funding this work.

References

[1] G. Arechavaleta, J.-P. Laumond, H. Hicheur, and A. Berthoz. An optimality principle governing human walking. *IEEE Transactions on Robotics*, 24(1):5–14, 2008.

[2] P. W. Fink, P. S. Foo, and W. H. Warren. Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception*, 4(1):2, 2007.

[3] K. Gidlöf, A. Wallin, R. Dewhurst, and K. Holmqvist. Using eye tracking to trace a cognitive process: Gaze behaviour during decision making in a natural environment. *Journal of Eye Movement Research*, 6(1):1–14, 2013.

[4] R. Grasso, P. Prévost, Y. P. Ivanenko, and A. Berthoz. Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. *Neuroscience Letters*, 253(2):115–118, 1998.

[5] H. Groenda, F. Nowak, P. Rößler, and U. D. Hanebeck. Telepresence techniques for controlling avatar motion in first person games. In *Intelligent Technologies for Interactive Entertainment*, pages 44–53. Springer, 2005.

[6] M. A. Hollands, A. Patla, and J. Vickers. look where youre going!: gaze behaviour associated with maintaining and changing the direction of locomotion. *Experimental Brain Research*, 143(2):221–230, 2002.

[7] V. Interrante, B. Ries, and L. Anderson. Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments. In *IEEE Symposium on 3D User Interfaces*. IEEE, 2007.

[8] A. Jain, H. S. Koppula, B. Raghavan, and A. Saxena. Know before you do: Anticipating maneuvers via learning temporal driving models. *arXiv preprint arXiv:1504.02789*, 2015.

[9] S. Koo and D.-S. Kwon. Recognizing human intentional actions from the relative movements between human and robot. In *The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 939–944. IEEE, 2009.

[10] M. F. Land. The visual control of steering. *Vision and action*, pages 163–180, 1998.

[11] M. F. Land. Eye movements and the control of actions in everyday life. *Progress in retinal and eye research*, 25(3):296–324, 2006.

[12] M. F. Land and D. N. Lee. Where do we look when we steer. *Nature*, 369(6483):742–744, 1994.

[13] D. R. Montello. *The Cambridge handbook of visuospatial thinking*, chapter Navigation, pages 257–294. Cambridge University Press, 2005.

[14] T. Nescher, Y.-Y. Huang, and A. Kunz. Planning redirection techniques for optimal free walking experience using model predictive control. *IEEE Symposium on 3D User Interfaces*, 2014.

[15] T. Nescher and A. Kunz. Using head tracking data for robust short term path prediction of human locomotion. In *Transactions on Computational Science XVIII: Special Issue on Cyberworlds*. Springer, 2013.

[16] N. Nitzsche, U. D. Hanebeck, and G. Schmidt. Motion compression for telepresent walking in large target environments. *Presence: Teleoperators and Virtual Environments*, 13(1):44–60, 2004.

[17] D. Novak, X. Omlin, R. Leins-Hess, and R. Riener. Predicting targets of human reaching motions using different sensing technologies. *IEEE Transactions on Biomedical Engineering*, 60(9):2645–2654, 2013.

[18] J. L. Orquin and S. M. Loose. Attention and choice: A review on eye movements in decision making. *Acta psychologica*, 144(1):190–206, 2013.

[19] T. Peck, M. Whitton, and H. Fuchs. Evaluation of reorientation techniques for walking in large virtual environments. In *VR '08: Proceedings of the IEEE Conference on Virtual Reality*, pages 121–127. IEEE, 2008.

[20] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected walking. In *Proceedings of EUROGRAPHICS*, volume 9, pages 105–106, 2001.

[21] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe. Task and context determine where you look. *Journal of vision*, 7(14), 2007.

[22] F. Steinicke, G. Bruder, L. Kohli, J. Jerald, and K. Hinrichs. Taxonomy and implementation of redirection techniques for ubiquitous passive haptic feedback. In *International Conference on Cyberworlds*, pages 217–223. IEEE, 2008.

[23] F. Steinicke, G. Bruder, T. Ropinski, and K. Hinrichs. Moving towards generally applicable redirected walking. In *Proceedings of the Virtual Reality International Conference*, pages 15–24, 2008.

[24] J. Su. Motion compression for telepresence locomotion. *Presence: Teleoperators and Virtual Environments*, 16(4):385–398, 2007.

[25] J. M. Wiener, O. De Condappa, and C. Hölscher. Do you have to look where you go? gaze behaviour during spatial decision making. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 1583–1588, 2011.

[26] J. M. Wiener, C. Hölscher, S. Büchner, and L. Konieczny. Gaze behaviour during space perception and spatial decision making. *Psychological research*, 76(6):713–729, 2012.

[27] B. Williams, G. Narasimham, T. P. McNamara, T. H. Carr, J. J. Rieser, and B. Bodenheimer. Updating orientation in large virtual environments using scaled translational gain. In *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 21–28. ACM, 2006.

[28] B. Williams, G. Narasimham, B. Rump, T. P. McNamara, T. H. Carr, J. Rieser, and B. Bodenheimer. Exploring large virtual environments with an hmd when physical space is limited. In *Proceedings of the 4th symposium on Applied perception in graphics and visualization*, pages 41–48. ACM, 2007.

[29] H. C. Yen, H. P. Huang, and S. Y. Chung. Goal-directed pedestrian model for long-term motion prediction with application to robot motion planning. In *IEEE Workshop on Advanced robotics and Its Social Impacts*, pages 1–6. IEEE, 2008.

[30] M. Zank and A. Kunz. Using locomotion models for estimating walking targets in immersive virtual environments. In *International Conference on Cyberworlds*, pages 229–236, 2015.

[31] M. Zmuda, J. L. Wonser, E. R. Bachmann, E. Hodgson, et al. Optimizing constrained-environment redirected walking instructions using search techniques. *IEEE Transactions on Visualization and Computer Graphics*, 19(11):1872–1884, 2013.